DOCUMENT RESUME

ED 385 549 TM 023 965

AUTHOR Mislevy, Robert J.
TITLE Probability-Based Inference in Cognitive Diagnosis.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
REPORT NO ETS-RR-94-3-ONR
PUB DATE Feb 94
CONTRACT N00014-88-K-0304
NOTE 60p.; Paper presented at the Office of Naval Research Contractors Conference (Iowa City, IA, May 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Bayesian Statistics; Cognitive Processes; Cognitive Psychology; *Educational Diagnosis; *Inferences; Knowledge Level; *Learning; Models; Networks; *Probability; Subtraction

ABSTRACT

Recent developments in cognitive psychology suggest models for knowledge and learning that often fall outside the realm of standard test theory. This paper concerns probability-based inference in terms of such models. The essential idea is to define a space of "student models"--simplified characterizations of students' knowledge, skill, or strategies, indexed by variables that signify their key aspects. From theory and data, one posits probabilities for the ways that students with different configurations in this space will solve problems, answer questions, and so on. Then the machinery of probability theory allows one to reason from observations of a student's actions to likely values of parameters in a student model. An approach using Bayesian inference networks is outlined. Basic ideas of structure and computation in inference networks are discussed and illustrated with an example from the domain of mixed-number subtraction. Six tables and 11 figures illustrate the discussion. (Contains 41 references.) (Author/SLD)

RR-94-3-ONR

PROBABILITY-BASED INFERENCE IN COGNITIVE DIAGNOSIS

Robert J. Mislevy

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator

Educational Testing Service
Princeton, NJ

February 1994

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1994	Final

4. TITLE AND SUBTITLE
Probability-Based Inference in Cognitive Diagnosis

5. FUNDING NUMBERS
G. N00014-88-K-0304
PE. 61153N
PR. RR 04204
TA. RR 04204-01
WU. R&T 4421552

6. AUTHOR(S)
Robert J. Mislevy

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-94-3-ONR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences
Code 1142CS
Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified/Unlimited

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT *(Maximum 200 words)*

Recent developments in cognitive psychology suggest models for knowledge and learning that often fall outside the realm of standard test theory. This paper concerns probability-based inference in terms of such models. An approach utilizing Bayesian inference networks is outlined. Basic ideas of structure and computation in inference networks are discussed, and illustrated with an example from the domain of mixed-number subtraction.

14. SUBJECT TERMS
Bayesian inference, belief nets, cognitive diagnosis, psychology, educational measurement, inference networks.

15. NUMBER OF PAGES
57

16. PRICE CODE
N/A

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassifed	Unclassified	Unclassified	SAR

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

Probability-Based Inference in Cognitive Diagnosis

Robert J. Mislevy

Educational Testing Service

March, 1994

Presented at the Office of Naval Research Contractors Conference in Iowa City, IA, May, 1993. This work was supported by Contract No. N00014-91-J-4101, R&T 4421573-01, from the Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research. I gratefully acknowledge the assistance of Duanli Yan, and discussions with Anne Béland, Drew Gitomer, Irvin Katz, Sung-Ho Kim, Paul Nichols, Kikumi Tatsuoka, and Tony Thompson.

Abstract

Recent developments in cognitive psychology suggest models for knowledge and learning that often fall outside the realm of standard test theory. This paper concerns probability-based inference in terms of such models. An approach utilizing Bayesian inference networks is outlined. Basic ideas of structure and computation in inference networks are discussed, and illustrated with an example from the domain of mixed-number subtraction.

Key words: Bayesian inference, belief nets, cognitive diagnosis, cognitive psychology, educational measurement, inference networks.

Introduction

The psychological paradigm emerging from cognitive psychology suggests new models for students' capabilities—a potentially powerful framework to plan instruction, evaluate progress, and provide feedback to students and teachers (Snow & Lohmann, 1989). As in traditional test theory, however, we face problems of inference: Just what kinds of things are to be said about students, by themselves or others? What evidence is needed to support such statements? How much faith can we place in the evidence, and in the statements? How do we sort out elements of evidence that are overlapping, redundant, or contradictory? When do we need to ask different questions or pose additional situations to distinguish among competing explanations of what we see?

This paper discusses a probabilistic framework for addressing questions like these. The essential idea is to define a space of "student models"—simplified characterizations of students' knowledge, skill, and/or strategies, indexed by variables that signify their key aspects. From theory and data, one posits probabilities for the ways that students with different configurations in this space will solve problems, answer questions, and so on. This done, the machinery of probability theory allows one to reason from observations of a student's actions to likely values of parameters in a student model.

Recent developments in statistical theory make it possible to carry out such inference in large and complex systems of variables. The program of research introduced here is beginning to explore the potential of this approach in educational assessment and cognitive diagnosis. By working out the details of specific illustrative examples, we are learning about the kinds of domains and student models that are practical to address, and starting to tackle an agenda of practical engineering challenges. We begin with an overview of inference networks, walking through a simple numerical example from medical diagnosis. An example from mixed-number subtraction illustrates the features of the approach as applied to cognitive assessment.

Probability-Based Inference

Inference is reasoning from what we know and what we observe to explanations, conclusions, or predictions. We are always reasoning in the presence of uncertainty. The information we work with is typically incomplete, inconclusive, amenable to more than one explanation. We attempt to establish the weight and coverage of evidence in what we observe. But the very first question we must address is "Evidence about what?" Schum (1987, p. 16) points out the crucial distinction between *data* and *evidence:* "A datum becomes evidence in some analytic problem when its *relevance* to one or more hypotheses being considered is established. ... [E]vidence is relevant on some hypothesis if it either increases or decreases the likeliness of the hypothesis. Without hypotheses, the relevance of no datum could be established." In educational assessment and cognitive diagnosis, we construct hypotheses around notions of the nature and acquisition of knowledge and skill.

Schum distinguishes three types of reasoning, the distinctions among which are central to this presentation. *Deductive reasoning* flows from generals to particulars, within an established framework of relationships among variables—from causes to effects, from diseases to symptoms, from the way a crime is committed to the evidence likely to be found at the scene, from a student's knowledge and skills to observable behavior. That is, under a given state of affairs, what are the likely outcomes? *Inductive reasoning* flows in the opposite direction, also within an established framework of relationships—from effects to possible causes, from symptoms to diseases, from observable behavior to probable configurations of a student's knowledge and skills. Given outcomes, what state of affairs led to them? In *abductive reasoning,* reasoning proceeds from observations to a *new* hypotheses, new variables, or new relationships among variables. "Such a 'bottom-up' process certainly appears similar to induction; but there is an argument that such reasoning is, in fact, different from induction since an existing hypothesis collection is enlarged in the process. Relevant evidentiary tests of this new hypothesis are then *deductively* inferred from the new hypothesis." (Schum, 1987, p.20).

The diagnostic approach discussed in this paper consists of a network of variables defining the student model space, the observable-outcome space, and the interrelationships among them. All three types of reasoning play a role:

- *Abductive* reasoning guides its construction, drawing upon research results and previous practice to suggest the basic structure and statistical analyses refine it. For example, Piaget (e.g., 1960) searched painstakingly for commonalities in the development of children's proportional reasoning abilities over years of unique learning episodes of individual children. Siegler's (1981) characterization of children's understandings of balance-beam problems as a sequence of increasingly sophisticated strategic flowcharts captures key aspects of some of these patterns, and provides a basis for a student model space (Mislevy, Yamamoto, & Anacker, 1992).

- *Deductive* reasoning, supplemented by parameter estimation, is used to posit distributions of observable variables given configurations of variables in the student model. In Siegler's study, this corresponds to determining how a child with a given set of strategies at her disposal might attack a given balance-beam problem, in terms of distributions of expected classes of actions.

- *Inductive* reasoning, embodied in the algebra of probability theory, guides reasoning from observations of a given student to inferences about her knowledge and skills, in terms of updated beliefs about student-model variables. This corresponds to characterizing our beliefs about which balance-beam strategies a child possesses after seeing her responses to a set of problems.

- *Abductive* reasoning, triggered by unexpected patterns in data, is called for again by the results of the inductive reasoning phase. Sometimes a particular child's responses will not be consistent with any of the student models in the simplified framework; inductive reasoning within this framework fails to provide a satisfactory working approximation of her knowledge and skill. In such cases, we need richer data to support further exploration, to generate new conjectures.

A key concept in probability-based inference is *conditional independence*: Defined generally, one subset of variables may be related in a population, but they are independent given the values of another subset of variables. In cognitive models, relationships among observations variables are "explained" by unobservable variables that characterize aspects of knowledge, skill, strategies, and so on. In Thompson's (1982) words, we ask "What can this person be thinking so that his actions make sense from his perspective?" or "What organization does the student have in mind so that his actions seem, to him, to form a coherent pattern?" Judah Pearl argues that creating such intervening variables is not merely a technical convenience, but a natural element in human reasoning:

> *"...conditional independence is not a grace of nature for which we must wait passively, but rather a psychological necessity which we satisfy actively by organizing our knowledge in a specific way. An important tool in such organization is the identification of intermediate variables that induce conditional independence among observables; if such variables are not in our vocabulary, we create them. In medical diagnosis, for instance, when some symptoms directly influence one another, the medical profession invents a name for that interaction (e.g., 'syndrome,' 'complication,' 'pathological state') and treats it as a new auxiliary variable that induces*

conditional independence; dependency between any two interacting systems is fully attributed to the dependencies of each on the auxiliary variable."

Pearl, 1988, p. 44.

Conditional independence is thus a conceptual tool to structure reasoning, helping to define variables, organize relationships, and guide deductive reasoning. In educational and psychological measurement, a heritage of statistical inference built around unobservable variables and induced conditional probability relationships extends back to Spearman's (e.g., 1907) early work with latent variables, to Wright's (1934) path analysis, to Lazarsfeld's (1950) latent class models. The resemblance of the inference networks presented below to LISREL diagrams (Jöreskog & Sörbom, 1989) is no accident! Our work shares inferential machinery with this tradition, but extends the universe of discourse to student models suggested by cognitive psychology.

Inference Networks

Probability-based inference in complex networks of interdependent variables is an active topic in statistical research, spurred by applications in such diverse areas as forecasting, pedigree analysis, troubleshooting, and medical diagnosis (e.g., Lauritzen & Spiegelhalter, 1988; Pearl, 1988). Current interest centers on obtaining the distributions of selected variables conditional on observed values of other variables, such as likely characteristics of offspring of selected animals given characteristics of their ancestors, or probabilities of disease states given symptoms and test results. As we shall see below, conditional independence relationships, as suggested by substantive theory, play a central role in the topology of the network of interrelationships in a system of variables. If the topology is favorable, such calculations can be carried out in real time in large systems by means of strictly local operations on small subsets of interrelated variables ("cliques") and their intersections.

This section briefly reviews basic concepts of construction and local computation for inference networks. Details can be found in the statistical and expert-systems literature; Lauritzen and Spiegelhalter (1988), Pearl (1988), and Shafer and Shenoy (1988), for example, discuss updating strategies, a kind of generalization of Bayes theorem. Computer programs are commercially available to carry out the number-crunching aspect. We used Andersen, Jensen, Olesen, and Jensen's (1989) HUGIN program and Noetic System's (1991) ERGO for the examples in this presentation.

To move from a structure of interrelationships among variables to a representation amenable to real-time local calculation, the steps listed below are taken. The first two encompass defining the key variables in an application and explicating their interrelationships. In essence, this information is the input to programs like ERGO and HUGIN, which then carry out Steps 3 through 7.

Step 1. Recursive representation of the joint distribution of variables.

Step 2. Directed graph representation of (1).

Step 3. Undirected, triangulated graph.

Step 4. Determination of cliques and clique intersections

Step 5. Join tree representation.

Step 6. Potential tables.

Step 7. Updating scheme.

Although computer programs are available, it is useful nevertheless to walk through the details of simple example—to watch what happens inside the "black box"—to develop intuition that can guide more ambitious applications. We borrow a simple example from Andreassen, Jensen, and Olesen (n.d.). It concerns two possible diseases a particular patient may have, flu and throat infection (FLU and THRINF), and two possible symptoms, fever and sore throat (FEV and SORETHR). The diseases are modeled as independent, and the symptoms as conditionally independent given disease states. These relationships are depicted in Figure 1, which will be discussed in greater detail below. All four variables can take values of "yes" and "no." We assume that exactly one value characterizes each variable for a patient, although we may not know these values with certainty. We employ probabilities to express our states of belief. We note in passing that it would be possible to work with the full joint distribution of the four variables in this example directly, using the textbook form of Bayes theorem to update beliefs of disease states as symptoms become known. This approach rapidly becomes infeasible as the number of variables in the system increases, whereas the approach described below has been employed in networks with over 1000 variables (Andreassen, Woldbye, Falck, & Andersen, 1987).

[[Figure 1 about here]]

1. A recursive representation of the joint distribution of variables

A *recursive representation* of the joint distribution of a set of random variables $X_1, \ldots, X_N$ takes the form

$$p(X_1,\ldots,X_n) = p(X_n|X_{n-1},\ldots,X_1)\, p(X_{n-1}|X_{n-2},\ldots,X_1)\cdots p(X_2|X_1)\, p(X_1)$$

$$= \prod_{j=1}^{n} p(X_j|X_{j-1},\ldots,X_1)\,, \qquad (1)$$

where the term for j=1 is defined as simply $p(X_1)$. A recursive representation can be written for any ordering of the variables, but one that exploits conditional independence relationships can prove more useful as variables drop out of the conditioning lists. This is where substantive theory comes into play; for example, modeling conditional probabilities of symptoms given disease states, rather than vice versa. The following representation exploits the independence of FLU and THRINF and the conditional independence of FEV and SORETHR:

P(FEV, SORETHR, FLU, THRINF)

= P(FEV | SORETHR, FLU, THRINF) P(SORETHR | FLU, THRINF) P(FLU | THRINF) P(THRINF)

= P(FEV | FLU, THRINF) P(SORETHR | FLU, THRINF) P(FLU) P(THRINF). (2)

Equation 2, like Figure 1, indicates the qualitative dependence structure of the relationships among the variables without specifying quantitative values. Constructing the full joint distribution from the recursive representation requires the specification of conditional probability distributions for each variable. For each combination of values of a variable's parents, this matrix gives the conditional probabilities of each of its potential values. Associated with variables having no parents, such as FLU and THRINF, is a vector of base rates or prior probabilities. We shall assign to both FLU and THRINF prior probabilities of .11 for "yes" and .89 for "no." This might correspond to base rates in a reference population to which our patient belongs. Conditional probabilities of FEV and of SORETHR given all combinations of FLU and THRINF appear in Table 1. In practice, such probabilities would be determined by disease theory, physiological principles, and past experience. The tabled values indicate that...

- Throat infection usually causes a sore throat whether or not flu is also present (.91 and .90 respectively); flu alone occasionally leads to a sore throat (.05), but the chances of a sore throat without either flu or throat infection is only .01.

- Having both flu and throat infection leads almost certainly to fever (.99); either disease by itself leads to fever with probability .90; and the probability of fever when neither disease is present is only .01.

[[Table 1 about here]]

The updating schemes discussed below assume these conditional probabilities are known with certainty. In practice, of course, they are usually not. Current research in the field includes characterizing the impact this source of uncertainty, sequentially improving estimates as additional data are obtained, and incorporating this uncertainty formally by augmenting the network with variables that parameterize the extent of knowledge about conditional probabilities (Spiegelhalter, 1989).

2. A directed graph representation of the joint distribution of variables

Corresponding to the algebraic representation of Equation 1 is a graphical representation—a *directed acyclic graph* (DAG). The graph inherits its "directedness" and "acyclic" properties from the recursive expression of the distribution in Equation 1. Direction comes from which variables are written as conditional on others in the representation, and the recursive expression prohibits "cycles" such as "A depends on B, B depends on C, and C depends on A." Figure 1, corresponding to Equation 2, is the DAG for our example. Each variable is a node in the graph; directed arrows run from "parents" to "children,"[1] indicating conditional dependence relationships among the variables.

A DAG depicts the qualitative structure of associations among variables in the domain. Theory about the domain is the starting point, but a real application requires model-fitting, model evaluation, and model refinement. While many standard techniques from statistical theory are useful in this endeavor, certain complications arise. In large networks, for example, many cases will be incomplete; there is no practical need to obtain the results of additional detailed diagnostic tests for diseases that have already been ruled out. And while global tests of model fit are useful in comparing alternative models, more focused tests checking local features of models and verifying predictions one case at a time are more useful for model refinement. While the updating schemes discussed below take the DAG structure as given, we must keep in mind that the success of an application ultimately depends on the care and thought that go into developing that structure.

3. An undirected, triangulated graph

Starting with the DAG, one drops the directions of the associations and adds edges as necessary to meet two requirements. First, the parents of a given child must be connected. Secondly, the graph must be *triangulated*; that is, any path of connections from a variable back to itself (a loop) consisting of four or more variables must have a chord, or "short cut." Triangulation is necessary for expressing probability relationships in a way that lends itself to coherent propagation of information. Kim and Pearl's (1983) initial work with individual variables showed how to carry out coherent local updating in *singly connected* networks of variables, or networks of variable associations with no loops at all. Most networks are not singly connected, however. Even our simple example has loops; for example, one can start a path at FEVER, follow a connection to FLU, then to SORTHR, then to THRINF, and finally return to FEVER.

The more recent updating schemes discussed here generalize Kim and Pearl's ideas by arranging variables into subsets called cliques, in a way such that the cliques form a singly-connected graph. Generalizations of Kim and Pearl's approach can then be applied at the level of cliques. Triangulating the original graph of variables guarantees that a singly-connected clique representation can be constructed (Jensen, 1988). A triangulation scheme is not necessarily unique, and various algorithms have been developed to construct triangulated graphs that support efficient calculation (e.g., Tarjan & Yannakakis, 1984). Figure 2 is the undirected, triangulated graph for our example.

[[Figure 2 about here]]

4. Determination of cliques and clique intersections

From the triangulated graph, one determines *cliques,* subsets of variables that are all linked pairwise to one another. Cliques overlap, with sets of overlapping variables called *clique intersections*. Cliques and clique intersections constitute the structure for local updating. Figure 3 shows the two cliques in our example, {FEVER, FLU, THRINF} and {FLU, THRINF, SORTHR}. The clique intersection is {FLU, THRINF}.

[[Figure 3 about here]]

Just as there can be multiple ways to produce a triangulated graph from a given DAG, there can be multiple ways to define cliques from a triangulated graph. Algorithms for determining a clique structure that supports efficient calculation are also a focus of research. The amount of computation grows roughly geometrically with clique size, as measured by the number of possible configurations of all values of all variables in a clique. A clique representation with many small cliques is therefore preferred to a representation with a few larger cliques. Strategies for increased efficiency at this stage include redefining variables, adding variables to break loops, and dropping associations when the consequences are benign.

5. Join tree representation

A join-tree representation depicts the singly-connected structure of cliques and clique intersections. This is the structure through which local updating flows. A join tree exhibits the *running intersection* property: If a variable appears in two cliques, it appears in all cliques and clique intersections in the single path connecting them. Figure 4 gives the join-tree for our example.

[[Figure 4 about here]]

6. Potential tables

Local calculation is carried out with tables that convey the joint distributions of variables within cliques, or *potential tables*. Similar tables for clique intersections are used to pass updating information from one clique to another. The potential tables in Table 2 indicate the initial status of the network for our example; that is, before specific knowledge of a particular individual's symptoms or disease states becomes known. For example, the potential table for Clique 1 is calculated using the prior probabilities of .11 for both flu and throat infection, the assumption that they are independent, and the conditional probabilities of sore throat for each flu/throat-infection combination.

[[Table 2 about here]]

The initial probability for fever can be obtained by marginalizing the potential table for Clique 1 with respect to flu and throat infection. This amounts to summing down the "FEVER: yes" column, yielding a value of .20. Similarly, the initial probability for sore throat is obtained by summing down the "SORTHR: yes" column in the potential table for Clique 2, yielding .11.

7. Local updating

Absorbing new evidence about a single variable is effected by re-adjusting the appropriate margin in a potential table that contains that variable, then propagating the resulting change to the clique to other cliques via the clique intersections. This process continues outward from the clique where the process began, until all cliques have been updated. The single-connectedness and running intersection properties of the join tree assure that coherent probabilities result.

Suppose that we learn the patient in our example *does* have a fever. How does this change our beliefs about the other variables? The calculations are summarized in Table 3.

[[Table 3 about here]]

- The process begins with the potential table for Clique 1. In the initial condition, we had a joint probability distribution for the variables in this clique, say, f_0(FEVER, FLU, THRINF). We now know with certainty that FEVER=yes, so the column for FEVER=no is zeroed out.[2] Denote the updated potential table f_1(FEVER, FLU, THRINF). One could re-normalize the entries in the FEVER=yes column at this point, but only the proportionality information needs to be sent on for updating.

- The clique intersection table is updated to reflect the new proportional relationships among the probabilities for FLU and THRINF, or f_1(FLU, THRINF). Normalizing them to sum to one would give probabilities, P_1(FLU, THRINF), which marginalize to .51 for FLU=yes and for THRINF=yes.

- The potential table for Clique 2 is updated by first dividing all entries in a row by the value for that row in the original clique intersection table, then multiplying them by the corresponding entries in the new one obtained in the previous step. The resulting entries are proportional to the new posterior probabilities for the variables in Clique 2. We now examine the rationale for this step in terms of probabilities (but recall that it suffices within the black box to simply pass the correct information about proportionalities along the join tree).

The initial joint probability distribution for Clique 2, P_0(FLU,THRINF,SORTHR), implied beliefs about flu and throat infection that were consistent with those in the initial status of Clique 1. But incoming information about fever modified belief about flu and throat infection, to P_1(FLU,THRINF). We want to revise the information in the potential table for Clique 2 so that it is (1) consistent with the new beliefs about flu and throat infection, but (2) unchanged in terms of the relationship of sore throat conditional on fever and throat infection. This is accomplished as shown below, justifying the divide-by-old-and-multiply-by-new algorithm:

$$
\begin{aligned}
&P_1(\text{FLU},\text{THRINF},\text{SORTHR}) \\
&\quad = P(\text{SORTHR}\mid\text{FLU},\text{THRINF})\; P_1(\text{FLU},\text{THRINF}) \\
&\quad = \left[\frac{P(\text{SORTHR}\mid\text{FLU},\text{THRINF})\; P_0(\text{FLU},\text{THRINF})}{P_0(\text{FLU},\text{THRINF})}\right] P_1(\text{FLU},\text{THRINF}) \\
&\quad = \left[\frac{P_0(\text{FLU},\text{THRINF},\text{SORTHR})}{P_0(\text{FLU},\text{THRINF})}\right] P_1(\text{FLU},\text{THRINF}).
\end{aligned}
$$

- The entries in the Clique 2 potential table can be re-normed to sum to one, as shown in the final panel in Table 3, to facilitate the calculation of individual combinations of values or of margins. For example, the revised probability for sore throat is .48.

Application to Cognitive Diagnosis

The approach we are exploring begins in a specific application by defining a universe of student models. This "supermodel" is indexed by parameters that signify distinctions between states of understanding. Symbolically, we shall refer to the (typically vector-valued) parameter of the student-model as η. A particular set of values of η specifies a particular student model, or one particular state among the universe of possible states the supermodel can accommodate. These parameters can be qualitative or quantitative, and qualitative parameters can be unordered, partially ordered, or completely ordered. A supermodel can contain any mixture of these types. Their nature is derived from the structure and the psychology of the learning area, with the goal of being able to express essential distinctions among states of knowledge and skill .

Any application faces a modeling problem, a task construction problem, and an inference problem.

The *modeling* problem is delineating the states or levels of understanding in a learning domain. In meaningful applications this might address several distinct strands of learning, as understanding develops in a number of key concepts, and it might address the connectivity among those concepts. This substep defines the *structure* of $p(\mathbf{x}|\eta)$, where $\mathbf{x}$ represents observations. An interesting special case occurs when the universe of student models can be expressed as performance models (Clancey, 1986). A performance model consists of a knowledge base and manipulation rules that can be run on problems in a domain of interest. A particular model can contain both knowledge and production rules that are incorrect or incomplete; the solutions it produces will be correct or incorrect in identifiable ways. Here the parameter η specifies features of performance models, such as the set of production rules that characterizes a student's state of competence.

Obviously any model will be a gross simplification of the reality of cognition. A first consideration in what to include in the supermodel is the substance and the psychology of the domain: Just what are the key concepts? What are important ways of understanding and misunderstanding them? What are typical paths to competence? A second consideration is the so-called grain-size problem, or the level of detail at which student-models should differ. A major factor in answering this question is the decision-making framework under which the modeling will take place. As Greeno (1976) points out, "It may not be critical to distinguish between models differing in processing details if the details lack important implications for quality of student performance in instructional situations, or the ability of students to progress to further stages of knowledge and understanding."

An analog for the student model space is Smith & Wesson's "Identikit," which helps police construct likenesses of suspects. Faces differ in infinitely many ways, and skilled police artists can sketch infinitely many drawings to match witnesses' recollections (which is not the say that police artists' drawings duplicate suspects' faces perfectly; uncertainty enters in the link through the witness). Departments that can't support an artist use an Identikit, a collection of various face shapes, noses, ears, hair styles, and so on, that can be combined to approximate witnesses' recollections from a large, but finite, range of possibilities. The payoff lies not in how accurately the Identikit composite depicts the suspect, but whether it aids the search enough to justify its use.

Research relevant to constructing student models has been carried out in a wide variety of fields, including cognitive psychology, the psychology of mathematics learning and science learning, and artificial intelligence (AI) work on student modeling. Cognitive scientists have suggested general structures such as "frames" or "schemas" that can serve as a basis for modeling understanding (e.g., Minsky, 1975; Rumelhart, 1980), and have begun to devise tasks that probe their features (e.g., Marshall, 1989, 1993). Researchers interested in the psychology of learning in subject areas such as proportional reasoning have focused on identifying key concepts, studying how they are typically acquired (e.g., in mechanics, Clement, 1982; in ratio and proportional reasoning, Karplus, Pulos, & Stage, 1983), and constructing observational settings that allow one to infer students' understanding (e.g., van den Heuvel, 1990; McDermott, 1984). Our approach can succeed only by building upon foundations of such research.

The *task construction* problem is devising situations for which students who differ in the parameter space are likely to behave in observably different ways. The conditional probabilities of behavior of different types given the unobservable state of the student are the *values* of $p(\mathbf{x}|\eta)$, which may in turn be modeled in terms of another set of parameters, say β, that have to be estimated. The $p(\mathbf{x}|\eta)$ values provide the basis for inferring back about the student state. An element in $\mathbf{x}$ could contain a right or wrong answer to a multiple-choice test item; it could instead be the problem-solving approach regardless of whether the answer is right or wrong, the quickness of a responding, a characteristic of a think-aloud protocol, or an expert's evaluation of a particular aspect of the performance. The effectiveness of a task is reflected in differences in conditional probabilities associated with different parameter configurations, so a task may be very useful in distinguishing among some aspects of student models but useless for distinguishing among others (Marshall, 1989).

The *inference* problem is reasoning from observations to student models. This is where the inference network and local computation come into play. The model-building and item construction steps define the relevant variables (the student-model variables η and the observable variables $\mathbf{x}$) and provide conditional probabilities. Let $p(\eta)$ represent expectations about η in a population of interest—possibly non-informative, possibly based on expert opinion or previous analyses. Together, $p(\eta)$ and $p(\mathbf{x}|\eta)$ imply our initial expectations for what we might observe from a student. Once we make actual observations, we can revise our probabilities through the network to draw inferences about η given $\mathbf{x}$, via $p(\eta|\mathbf{x}) \propto p(\mathbf{x}|\eta)\ p(\eta)$. Thus $p(\eta|\mathbf{x})$ characterizes belief about a particular student's model after having observed a sample of the student's behavior.

Example: Mixed-Number Subtraction

This example illustrates a model that is aimed at the level of short-term instructional guidance. The form of the evidence being collected is traditional—right or wrong responses to open-ended mixed-number subtraction problems—but inferences are carried out in a student model motivated by cognitive analyses of the domain. It concerns which of two strategies students apply to the problems, and whether they are able to carry out procedures required singly or in combination in problems. Although a much finer grain-size can be entertained for models of these types of skills (e.g., VanLehn's 1990 analysis of whole number subtraction), this example incorporates the fact that whether an item is easy or hard to a given student depends in part on the strategy she employs. Rather than being discarded as noise, as it would be under standard test theory, this interaction is exploited by the analytic model as a source of evidence about a student's strategy usage.

The data and the cognitive analysis upon which the student model is grounded are due to Kikumi Tatsuoka (1987, 1990). The middle-school students she studied characteristically solve mixed number subtraction problems using one of two strategies:

Method A: Convert all whole and mixed numbers to improper fractions, subtract, then reduce if necessary.

Method B: Separate mixed numbers into whole number and fractional parts, subtract as two subproblems, borrowing one from minuend whole number if necessary, then reduce if necessary.

We analyzed 530 students' responses to 15 items. Table 4 shows how we characterized each item in terms of which of seven sub procedures would be required if it were solved with Method A and which would be required if it were solved with Method B.

The student model is comprised of a variable for which strategy a student uses and which of the subprocedures he is able to apply. The structure connecting the unobservable parameters of the student model and the observable responses is that ideally, a student using Method X (A or B, as appropriate to that student) would correctly answer items that under that strategy require only subprocedures the student has at his disposal (see Falmagne, 1989, Tatsuoka, 1990, and Haertel & Wiley, 1993, on models of this type). However, sometimes students miss items even under these conditions (false negatives), and sometimes they correctly answer items when they don't possess the subprocedures by other, possibly incorrect, means (false positives). The connection between observations and student model variables is thus probabilistic rather than deterministic.

[[Table 4 about here]]

A network for Method B

Figure 5 is a graphic depiction of the structural relationships in an inference network for Method B only. Nodes represent variables, and arrows represent dependence relationships. The joint probability distribution of all variables can be represented as the product of conditional probabilities, with each variable expressed in terms of conditional probabilities given its "parents." Five nodes, "Skill1" through "Skill5," represent basic subprocedures that a student who uses Method B might need use to solve items. Additional nodes, such as "Skills1&2" are conjunctions, representing, for example, either having or not having both Skill 1 and Skill 2. The node MN stands for "mixed number skills." It subsumes both Skill3, separating whole numbers from fractions, and Skill4, borrowing a unit from a whole number; the MN node contains the logical relationship that Skill3 is a prerequisite for Skill4. All of these skill variables and their combinations are represented in Figure 5 by rectangles. They are the elements of the student model, or η. The relationships among the skill nodes are either empirical (probabilities of having, say, Skill 2 given that one does or does not have Skill 1) or logical (one has "Skills1&2" only if one has both Skill 1 and Skill 2).

[[Figure 5 about here]]

The observables, **x**, are the actual test items. The ovular nodes representing items are children of nodes that represent the minimal necessary conjunction of skills necessary to solve that item if one uses Method B. The relationship between such a node and an item is probabilistic, indicating false positive and false negative probabilities.

Cognitive theory inspired the *structure* of this network. Initial estimates of the *numerical values* of conditional probability relationships were approximated using results from Tatsuoka's (1983) "rule space" of the data, with only students she classified as Method B users. That is, Dr. Tatsuoka's estimate of whether a student did or did not possess Skill1 and Skill2 were taken as truth, and our probabilities of students having Skill1, of having Skill2 given that they did or didn't have Skill2, and so on, are empirical proportions from this data set. (Duanli Yan and I are exploring the estimation of these conditional probabilities using the EM algorithm of Dempster, Laird, & Rubin, 1977.) Table 5 gives three examples of the conditional probabilities matrices we used as input to HUGIN and ERGO:

- Skill2 given Skill1. These are the conditional probabilities of having or not having Skill2, given that a student does or does not have Skill1. These were approximated from the results of Dr. Tatsuoka's analysis, as described above.

- Skills1&2 given Skill1 and Skill2. This is a logical relationship, indicating that a student has the conjunction of Skills 1 and 2 if and only if she has both Skill1 and Skill2.

- Item12 given Skills1&2. This matrix gives the probabilities of correctly answering Item 12, given that a student does or does not have the requisite set of skills under Method B. For the row in which Skills1&2 is true, we have the true positive and false negative success rates, .895 and .105 respectively. For the row in which Skills1&2 is false, we have the false positive and true negative rates, .452 and .548. (A relatively high false positive rate such as this often occur when an item on a test has appeared as a textbook example or homework exercise.)

[[Table 5 about here]]

Figure 6 presents the join tree for the DAG depicted in Figure 5. Figure 7 depicts base rate probabilities of skill possession and item percents-correct in the network with empirical associations, using the conditional probabilities from Tatsuoka's Rule Space analysis. This represents the state of knowledge one has about a student knowing that she uses Method B, but without having observed any item responses. Figure 8 shows how beliefs are changed after observing mostly correct answers to items requiring subprocedures other than Skill 2, but missing most of those that do require it. The base-rate and the updated probabilities for the five skills shown in Table 6 show substantial shifts toward the belief that the student commands Skills 1, 3, 4, and possibly 5, but almost certainly not Skill 2.

[[Figures 6-8 about here]]

[[Table 6 about here]]

A simplified network for Method B

An alternative representation exemplifies the tradeoffs one faces when building more complex networks, and illustrates their relationships to the network building and manipulation steps discussed above. A simpler network results if empirical relationships among skills are deleted, as shown in Figure 9. The resulting join tree is shown in Figure 10. The advantage of this simpler network is a join-tree with smaller maximally-sized clique, containing 4 variables rather than 5. The largest potential table has only 16 entries, rather than 48. By such simplifications, larger networks of variables can be updated in the same amount of calculating time. The simpler network uses only *direct* information from item responses to update beliefs about skill possession; that is, belief for Skill 3 is changed only by responses to items that require Skill 3. The tradeoff is the forfeiture of *indirect* information. Suppose we have ascertained that students who possess Skills 1 and 2 usually also possess Skill 3. The full network, incorporating this link, would revise our belief about Skill 3 in response to indirect evidence in the form of correct answers to items requiring Skills 1 and 2. The simplified network, omitting the link, would not revise belief about Skill 3 without direct evidence, or responses to items requiring Skill 3 itself.

[[Figures 9 & 10 about here]]

What kinds of inferential errors result from this simplification? Closed-form results with simple models indicate that ignoring positive relationships among unobservable variables higher in the network can lead to weaker, or more conservative, revision of belief about them from observations. This may be an acceptable price in some cases in return for being able to incorporate more variables into a network. (On the other hand, ignoring dependencies among observable variables can lead to overly strong updating—generally a more costly error.) A second rationale for omitting the empirical relationships among skills is that the resulting model, while conservative for a given population, may be more transportable to other populations—for example, students who studied fractions under a different curriculum. While the skill requirements of items may be fairly consistent over students, the relationships among skills may depend more heavily on the order and intensity with which they are studied.

A simultaneous network for both methods

We built a similar network for Method A. Figure 11 incorporates it and the Method B network into a single network that is appropriate when we don't know which strategy a student is using. Each item now has three parents: minimally sufficient sets of subprocedures under Method A and under Method B, and the new node "Is the student using Method A or Method B?" An item like $7\frac{2}{3} - 5\frac{1}{3}$ is hard under Method A but easy under Method B. An item like $2\frac{1}{3} - 1\frac{2}{3}$ is just the opposite. A response vector with most of the first type of items right and the second types wrong shifts belief toward the use of Method B, while the opposite pattern shifts belief toward the use of Method A. A pattern with mostly wrong answers gives posterior probabilities for Method A and Method B that are about the same as the base rate, but low probabilities for possessing any of the skills. We haven't learned much about which strategy such a student is using, but we do have evidence that he probably doesn't have subprocedure skills. Similarly, a pattern with mostly right answers again gives posterior probabilities for Method A and Method B that are about the same as the base rate, but high probabilities for possessing all of the skills. In any of these cases, the results could be used to guide an instructional decision.

[[Figure 11 about here--network for both methods]]

Extensions

This example could be extended in many ways, both as to the nature of the observations and the nature of the student model. With the present student model, one might explore additional sources of evidence about strategy use: monitoring response times, tracing solution steps, or simply asking the students to describe their solutions!

Each has tradeoffs in terms of cost and evidential value, and each could be sensible in some applications but not others. An important extension of the student model would be to allow for strategy switching (Kyllonen, Lohman, & Snow, 1984). Adults, for example, often decide whether to use Method A or Method B for a given item only after gauging which would be easier to apply. The variables in this more complex student model would express the tendencies of a student to employ various strategies under various conditions; students would then be mixtures in and of themselves, with "always use Method A" and "always use Method B" as extreme cases. Mixture problems are notoriously hard statistical problems; carrying out inference in the context of this more ambitious student model would certainly require the richer information mentioned above. Anne Béland and I (Béland & Mislevy, 1992) tackled this problem in the domain of proportional reasoning, addressing students' solutions to balance-beam tasks. We modeled students in terms of neo-Piagetian developmental stages based on the availability of certain concepts that could be fashioned into strategies for different kinds of tasks. The data for inferring a students' stages were their solutions and their explanations of the strategies they employed.

Conclusion

Inference network models can play useful roles in educational assessment. One is the use mentioned in our example, namely, cognitive diagnosis for short term instructional guidance as in an intelligent tutoring system (ITS). At ETS, we are currently working to implement probability-based inference updating the student model in an aircraft hydraulics ITS (Gitomer, Steinberg, & Mislevy, in press). Another is mapping out the evidential structure of observations and student knowledge structures (Haertel, 1989; Haertel & Wiley, 1993). As both models and observational contexts become more complex, more careful thought is required to sort out and characterize the implications and qualities of assessment tasks if we are to use the information effectively. We plan to explore the kinds of problems in which the approach outlined above proves efficacious, and to develop exemplars and methodological tools for employing it.

Notes

[1] This terminology is from the use of DAGs in pedigree analysis, where nodes represent characteristics of animals that are in fact parents and children.

[2] Partial information, such as "based on a reading from an unreliable thermometer, I'd place the probability of fever is .80," would lead to proportional re-adjustment of the columns, maintaining the proportional relationships *within* columns.

References

Andersen, S.K., Jensen, F.V., Olesen, K.G., & Jensen, F. (1989). *HUGIN: A shell for building Bayesian belief universes for expert systems* [computer program]. Aalborg, Denmark: HUGIN Expert Ltd.

Andreassen, S., Woldbye, M., Falck, B., & Andersen, S.K. (1987). MUNIN: A causal probabilistic network for interpretation of electromyographic findings. *Proceedings of the 10th International Joint Conference on Artificial Intelligence* (pp. 366-372). Milan: Kaufmann.

Béland, A., & Mislevy, R.J. (1992). Probability-based inference in a domain of proportional reasoning tasks. *ETS Research Report 92-15-ONR*. Princeton, NJ: Educational Testing Service.

Clement, J. (1982). Students' preconceptions of introductory mechanics. *American Journal of Physics, 50*, 66-71.

Dempster, A.P, Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Falmagne, J-C. (1989). A latent trait model via a stochastic learning theory for a knowledge space. *Psychometrika, 54*, 283-303.

Gitomer, D.H., Steinberg, L.S., & Mislevy, R.J. (in press). Diagnostic assessment of trouble-shooting skill in an intelligent tutoring system. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment*. Hillsdale, NJ: Erlbaum.

Greeno, J.G. (1976). Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In D. Klahr (Ed.), *Cognition and instruction* (pp. 123-159). Hillsdale, NJ: Erlbaum.

Haertel, E.H. (1989). Using restricted latent class models to map the skill structure of achievement test items. *Journal of Educational Measurement, 26*, 301-321.

Haertel, E.H., & Wiley, D.E. (1993). Representations of ability structures: Implications for testing. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 359-384). Hillsdale, NJ: Erlbaum.

Jensen, F.V. (1988). Junction trees and decomposable hypergraphs. Aalborg, Denmark: JUDEX Data Systems.

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: User's Reference Guide.* Mooresville, IN: Scientific Software, Inc.

Karplus, R., Pulos, S., & Stage, E. (1983). Proportional reasoning of early adolescents. In R.A. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes* (pp. 45-90). Orlando, FL: Academic Press.

Kim, J.H., & Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. *Proceedings of the 8th International Joint conference on Artificial Intelligence,* Karlsruhe, Germany (pp. 190-193).

Kyllonen, P.C., Lohman, D.F., & Snow, R.E. (1984). Effects of aptitudes, strategy training, and test facets on spatial task performance. *Journal of Educational Psychology, 76*, 130-145.

Lauritzen, S.L., & Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B, 50*, 157-224.

Lazarsfeld, P.F. (1950). The logical and mathematical foundation of latent structure analysis. In S.A. Stouffer, L. Guttman, E.A. Suchman, P.F. Lazarsfeld, S.A. Star, & J.A. Clausen, *Studies in social psychology in World War II, Volume 4: Measurement and prediction* (pp. 362-412). Princeton, NJ: Princeton university Press.

Marshall, S.P. (1989). Generating good items for diagnostic tests. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 433-452). Hillsdale, NJ: Erlbaum.

Marshall, S.P. (1993). Assessing schema knowledge. In N. Frederiksen, R.J. Mislevy, & I.I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 155-180). Hillsdale, NJ: Erlbaum.

Martin, J.D., & VanLehn, K. (in press). OLEA: Progress toward a multi-activity, Bayesian student modeler. *Proceedings of the International Conference on Artificial Intelligence in Education.*

McDermott, L.C. (1984). Research on conceptual understanding in mechanics. *Physics Today, 37*, 1-10.

Minsky, M. (1975). A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision* (pp. 211-277). New York: McGraw-Hill.

Mislevy, R.J., Yamamoto, K, & Anacker, S. (1992). Toward a test theory for assessing student understanding. In R.A. Lesh & S. Lamon (Eds.), *Assessments of authentic performance in school mathematics* (pp. 293-318). Washington, DC: American Association for the Advancement of Science.

Noetic Systems, Inc. (1991). ERGO [computer program]. Baltimore, MD: Author.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Kaufmann.

Piaget, J. (1960) The general problems of the psychological development of the child. In J.M. Tanner and B. Inhelder (Eds.), *Discussions on Child Development: Vol. 4. The fourth meeting of the World Health Organization Study Group on the Psychological Development of the Child, Geneva, 1956* (pp. 3-27). Geneva: World Health Organization.

Rumelhart, D.A. (1980). Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, & W. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.

Schum, D.A. (1987). *Evidence and inference for the intelligence analyst*. Lanham, Md.: University Press of America.

Shafer, G., & Shenoy, P. (1988). Bayesian and belief-function propagation. *Working Paper 121*. Lawrence, KS: School of Business, University of Kansas.

Siegler, R.S. (1981). Developmental sequences within and between concepts. *Monograph of the Society for Research in Child Development, Serial No. 189, 46.*

Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational measurement* (3rd Ed.) (pp. 263-331). New York: American Council on Education/Macmillan.

Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology, 18*, 161-169.

Spiegelhalter, D.J. (1986). Probabilistic reasoning in predictive expert systems. In L.W. Kanal and J. Lemmer (Eds.), *Artificial intelligence and statistics* (pp. 47-68). Amsterdam: North-Holland.

Tarjan, R.E., & Yannakakis, M. (1984). Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM Journal of Computation, 13*, 566-579.

Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement, 20*, 345-354.

Tatsuoka, K.K. (1987). Validation of cognitive sensitivity for item response curves. *Journal of Educational Measurement, 24*, 233-245.

Tatsuoka, K.K. (1990). Toward an integration of item response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M.G. Shafto, (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.

Thompson, P.W. (1982). Were lions to speak, we wouldn't understand. *Journal of Mathematical Behavior, 3*, 147-165.

van den Heuvel, M. (1990). Realistic arithmetic/mathematics instruction and tests. In K. Gravemeijer, M. van den Heuvel, & L. Streefland (Eds.), *Context free productions tests and geometry in realistic mathematics education* (pp. 53-78). Utrecht, The Netherlands: Research Group for Mathematical Education and Educational computer Center, State University of Utrecht.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions.* Cambridge, MA: MIT Press.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics, 5*, 161-215.

Table 1

Conditional Probabilities of Symptoms Given Disease States

FLU	THRINF	P(SORETHR)=yes	P(SORETHR)=no
yes	yes	.91	.09
yes	no	.05	.95
no	yes	.90	.10
no	no	.01	.99

FLU	THRINF	P(FEV)=yes	P(FEV)=no
yes	yes	.99	.01
yes	no	.90	.10
no	yes	.90	.10
no	no	.01	.99

Table 2

Potential Tables for Initial Status of Knowledge

Clique 1

FLU	THRINF	FEVER: yes	FEVER: no
yes	yes	.012	.000
yes	no	.088	.010
no	yes	.088	.010
no	no	.008	.784

FLU	THRINF	Probability
yes	yes	.012
yes	no	.098
no	yes	.098
no	no	.792

Clique 2

FLU	THRINF	SORTHR: yes	SORTHR: no
yes	yes	.011	.001
yes	no	.005	.093
no	yes	.088	.010
no	no	.008	.784

Table 3

Potential Tables af .r "FEVER=yes"

Clique 1

FLU	THRINF	FEVER: yes	FEVER: no
yes	yes	.012	0
yes	no	.088	0
no	yes	.088	0
no	no	.008	0

FLU	THRINF	Probability
yes	yes	.012
yes	no	.088
no	yes	.088
no	no	.008

Clique 2

FLU	THRINF	SORTHR: yes	SORTHR: no
yes	yes	.011	.001
yes	no	.004	.084
no	yes	.080	.009
no	no	.000	.008

Re-Normed Table for Clique 2

FLU	THRINF	SORTHR: yes	SORTHR: no
yes	yes	.059	.005
yes	no	.020	.426
no	yes	.406	.046
no	no	.000	.041

Table 4

Skill Requirements for Fractions Items

			If Method A used				If Method B used			
Item #	Text	1	2	5	6	7	2	3	4	5
4	$3\frac{1}{2}-2\frac{3}{2}=$	x			x		x	x	x	
6	$\frac{6}{7}-\frac{4}{7}=$	x								
7	$3-2\frac{1}{5}=$	x		x	x		x	x	x	x
8	$\frac{3}{4}-\frac{3}{8}=$	x								
9	$3\frac{7}{8}-2=$	x	x	x	x	x		x		
10	$4\frac{4}{12}-2\frac{7}{12}=$	x	x		x		x	x	x	
11	$4\frac{1}{3}-2\frac{4}{3}=$	x	x		x		x	x	x	
12	$\frac{11}{8}-\frac{1}{8}=$	x	x				x			
14	$3\frac{4}{5}-3\frac{2}{5}=$	x			x			x		
15	$2-\frac{1}{3}=$	x	x	x				x	x	x
16	$4\frac{5}{7}-1\frac{4}{7}=$	x	x		x			x		
17	$7\frac{3}{5}-\frac{4}{5}=$	x	x		x			x	x	
18	$4\frac{1}{10}-2\frac{8}{10}=$	x	x		x	x	x	x	x	
19	$7-1\frac{4}{3}=$	x	x	x	x	x	x	x	x	x
20	$4\frac{1}{3}-1\frac{5}{3}=$	x	x		x	x	x	x	x	

Skills:

1. Basic fraction subtraction
2. Simplify/Reduce
3. Separate whole number from fraction
4. Borrow one from whole number to fraction
5. Convert whole number to fraction
6. Convert mixed number to fraction
7. Column borrow in subtraction

Table 5

Examples of Conditional Probability Matrices for Method B Network

Skill2 given Skill1	Skill 2 Probabilities	
Skill 1 Status	Yes	No
Yes	.662	.338
No	.289	.711

Skills1&2 given Skill1, Skill2		Skills1&2 Probabilities	
Skill 1 Status	Skill 2 Status	Yes	No
Yes	Yes	1	0
Yes	No	0	1
No	Yes	0	1
No	No	0	1

Item12 given Skills1&2	Item 12 Probabilities	
Skills1&2 Status	Correct	Incorrect
Yes	.895	.105
No	.452	.548

Table 6

Prior and Posterior Probabilities of Subprocedure Profile

Skill(s)	Prior Probability	Posterior Probability
1	.883	.999
2	.618	.056
3	.937	.995
4	.406	.702
5	.355	.561
1 & 2	.585	.056
1 & 3	.853	.994
1, 3, & 4	.392	.702
1, 2, 3, & 4	.335	.007
1, 3, 4, & 5	.223	.492
1, 2, 3, 4, & 5	.200	.003

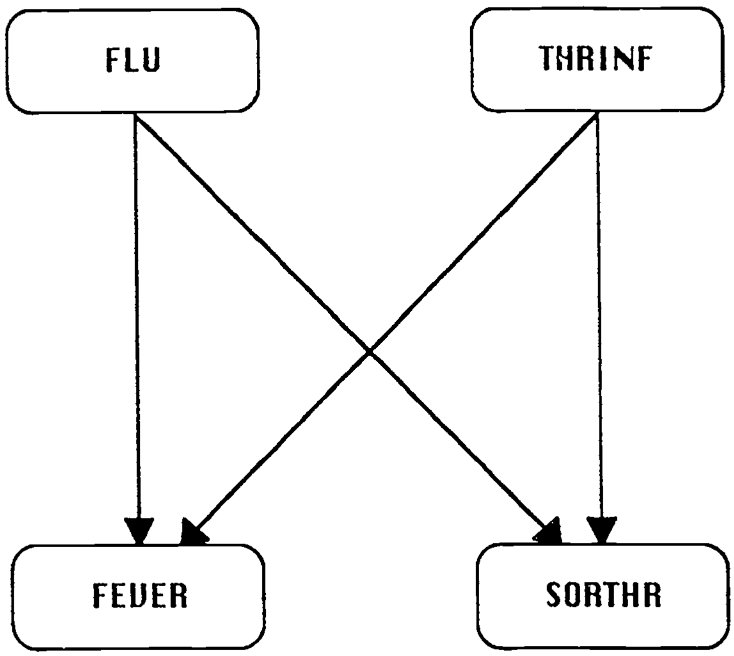


Figure 1

Directed Acyclic Graph Representation

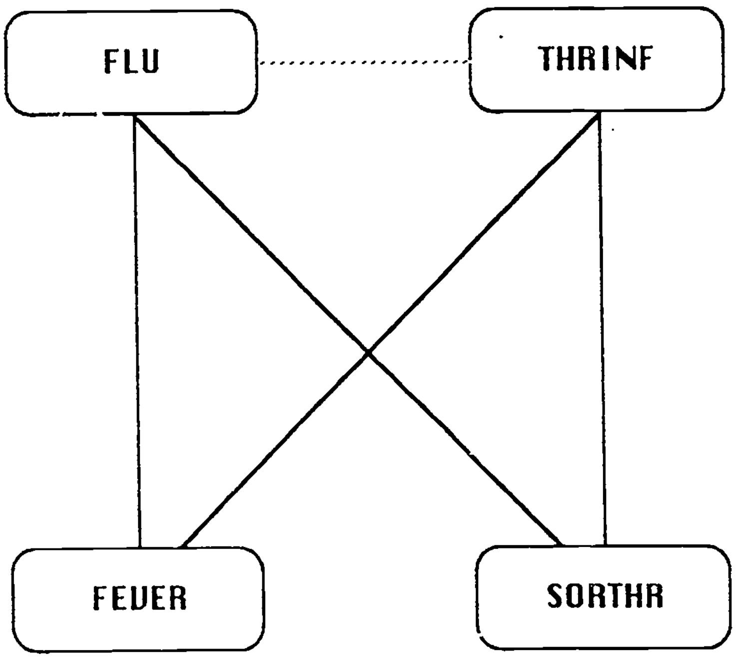
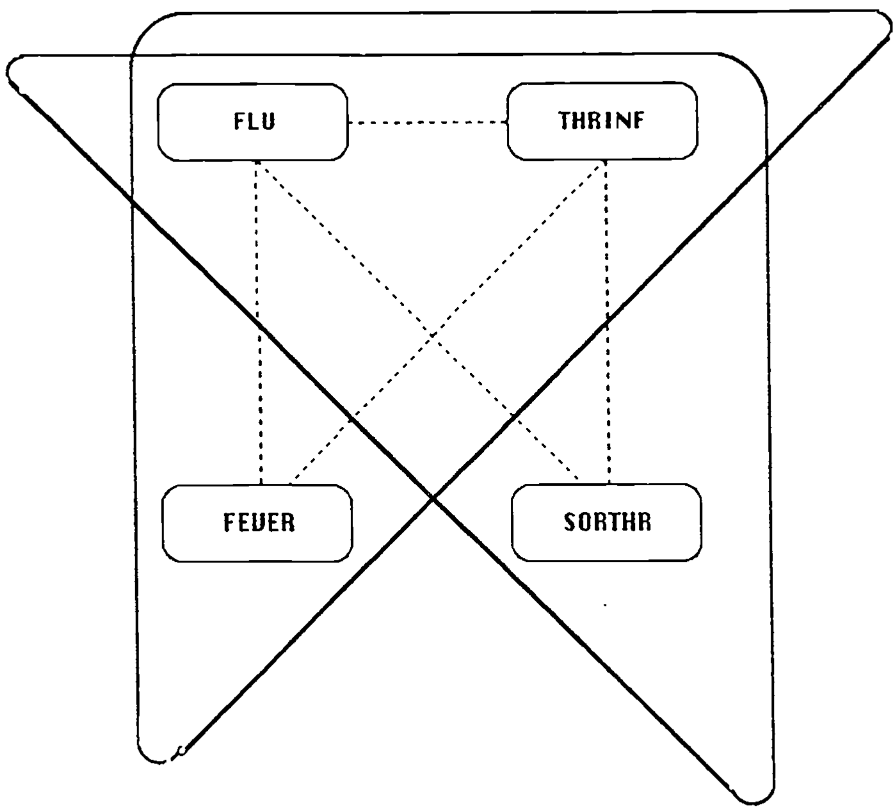


Figure 2

Undirected, Triangulated Graph Representation



Clique 1: FEVER, FLU, THRINF

Clique 2: FLU, THRINF, SORTHR

Figure 3

Clique Structure

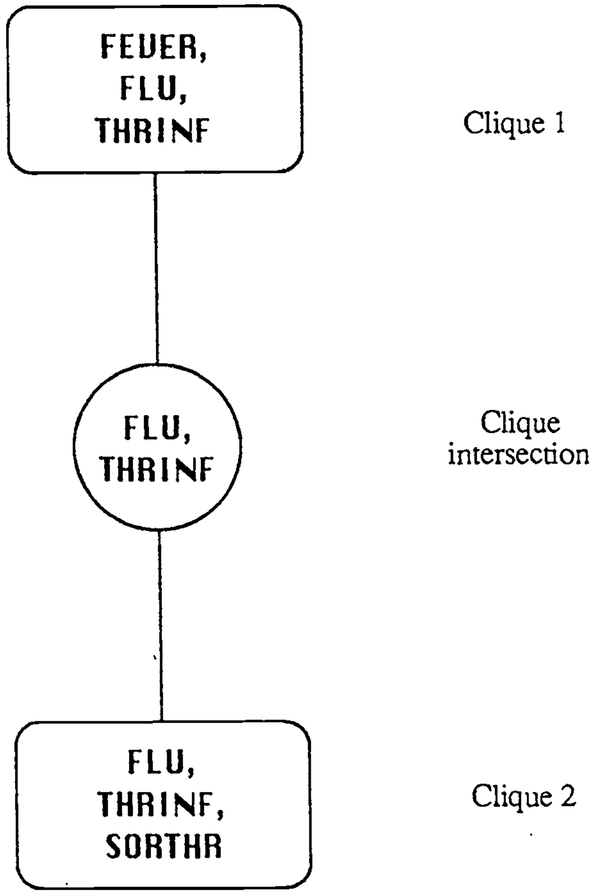


Figure 4

Join Tree Representation

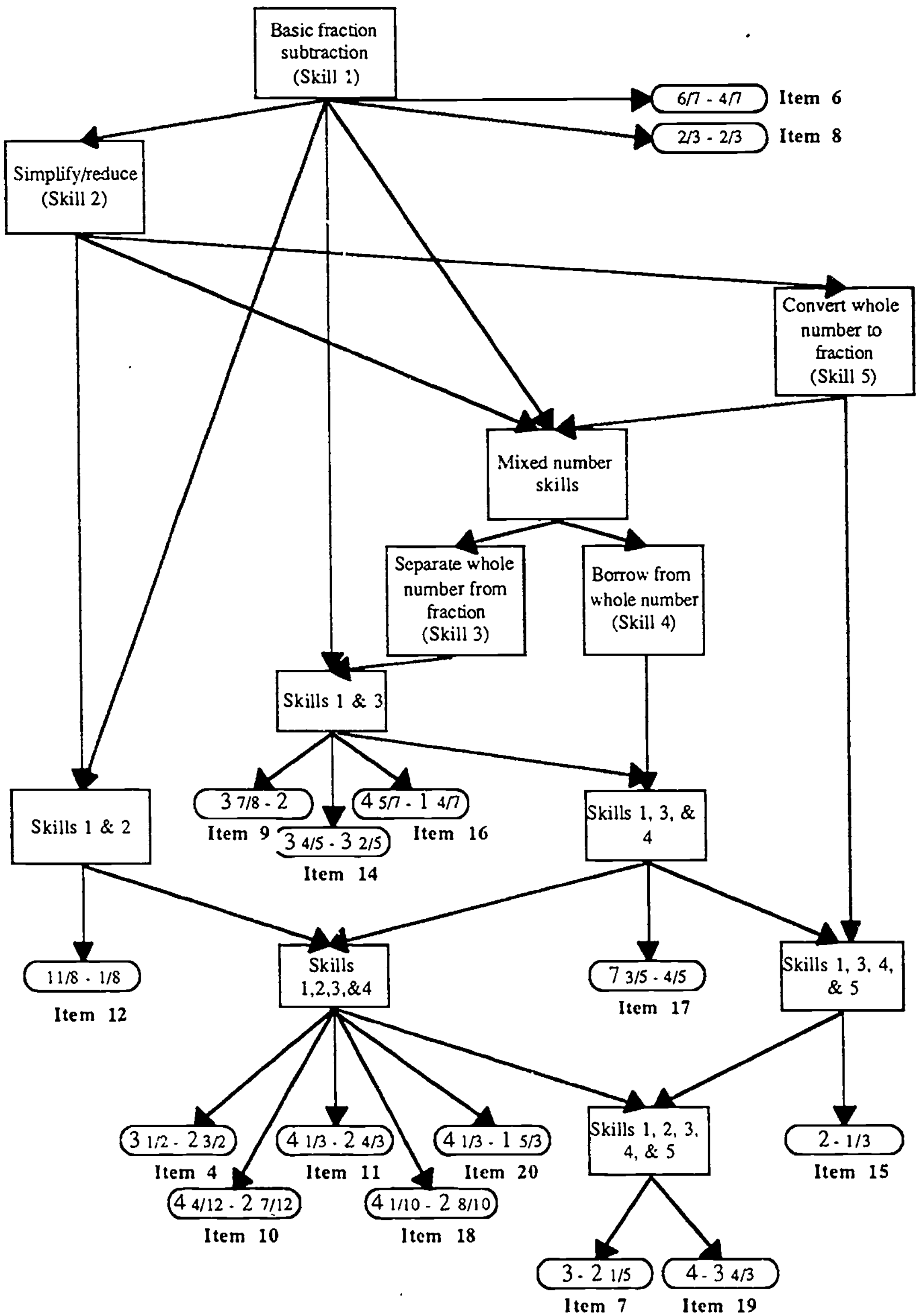


Figure 5

An Inference Network for Method B

BEST COPY AVAILABLE

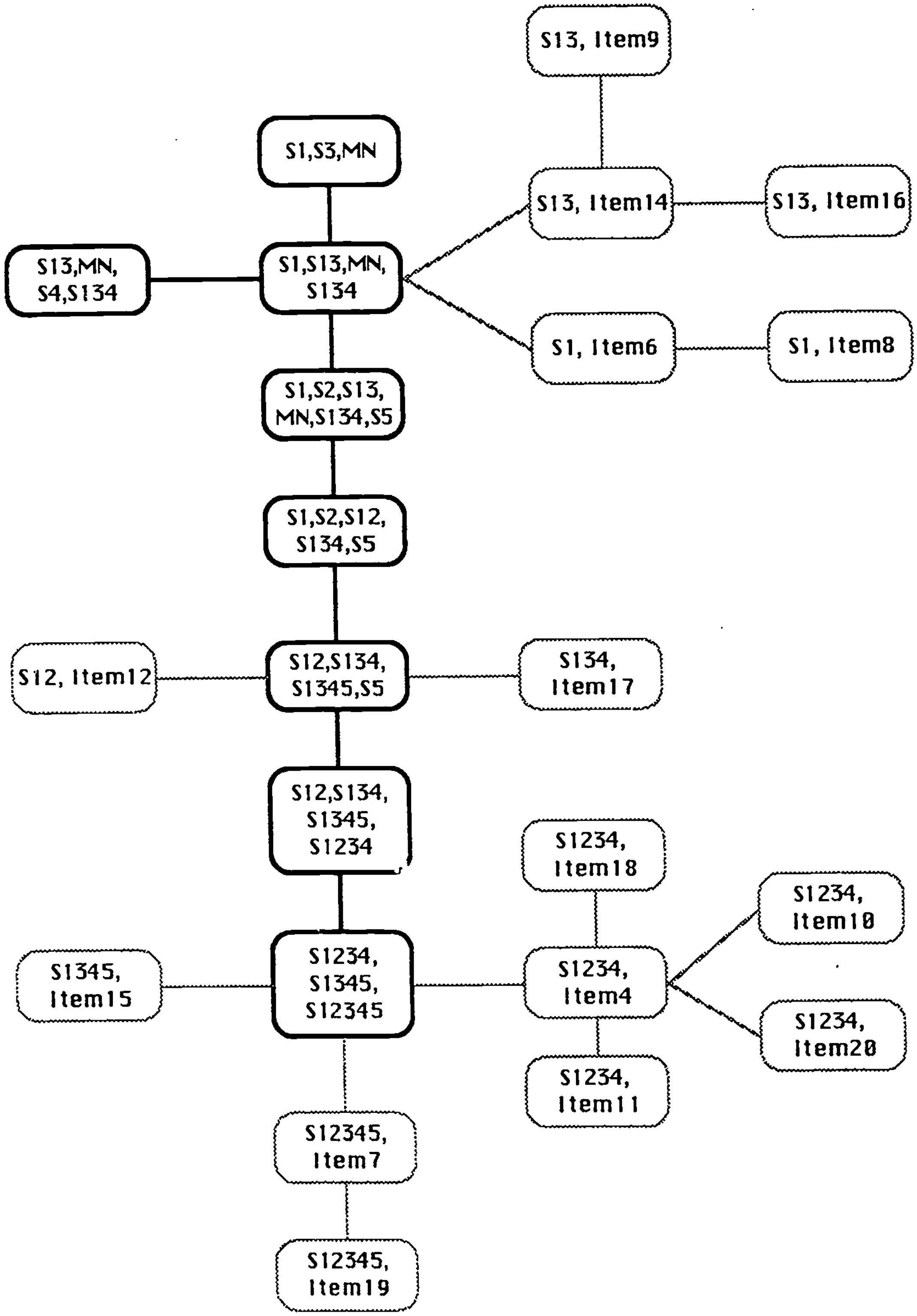
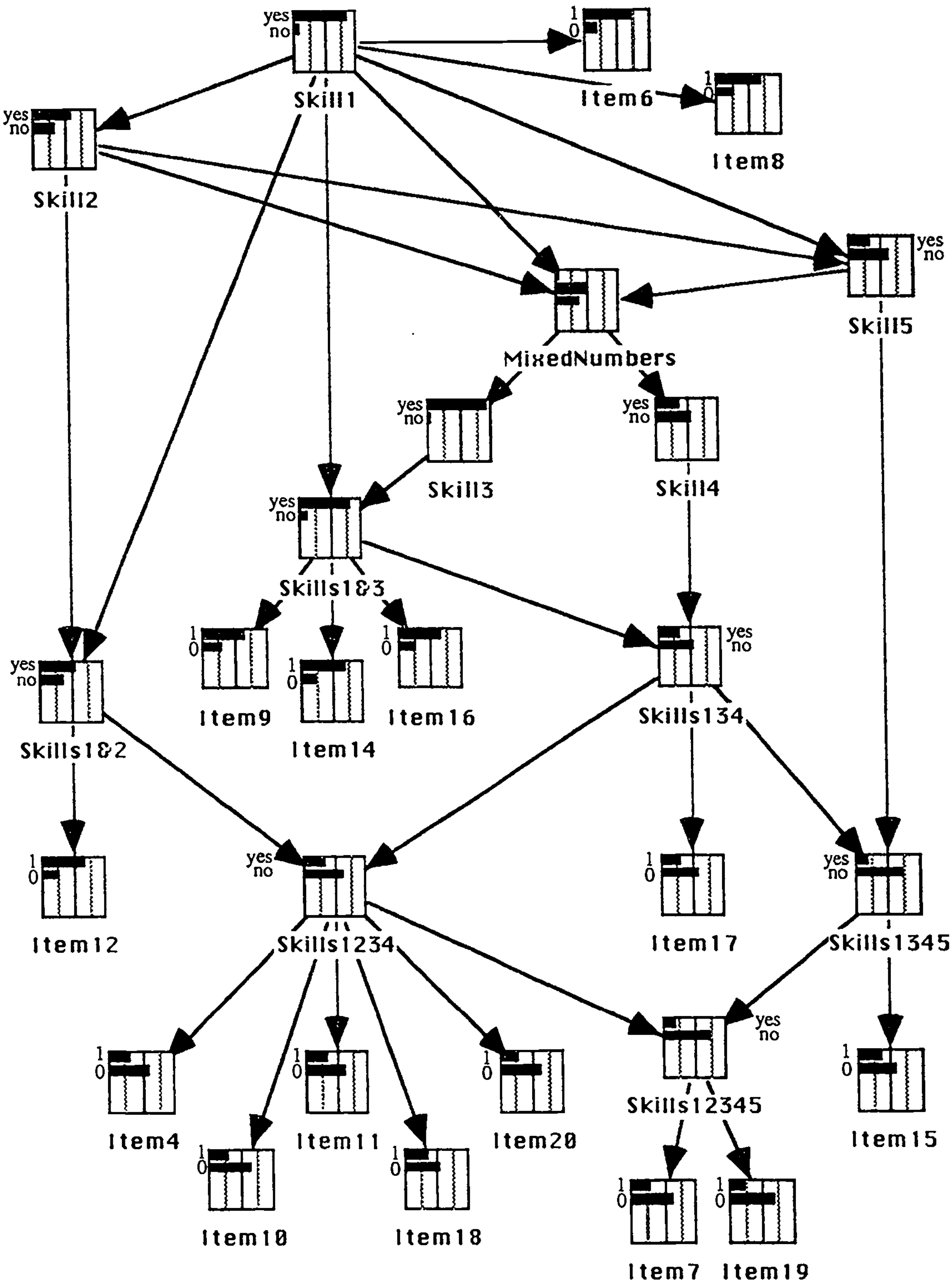


Figure 6

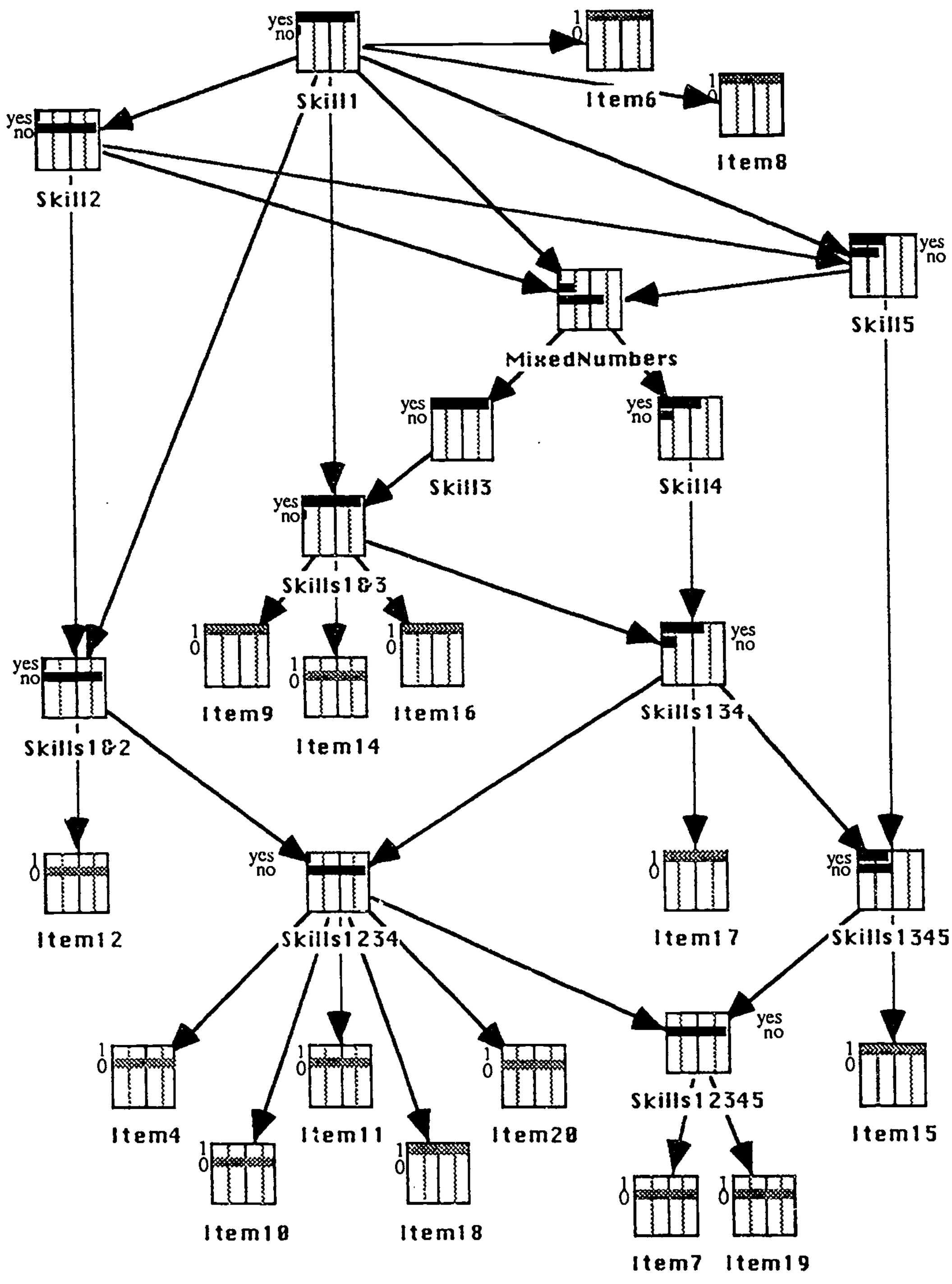
Join Tree for Network with Empirical Connections Among Skills



Note: Bars represent probabilities, summing to one for all the possible values of a variable.

Figure 7

Initial Probabilities for Method B Network



Note: Bars represent probabilities, summing to one for all the possible values of a variable. A shaded bar extending to one represents certainty, due to having observed value of that variable.

Figure 8

Posterior Probabilities for Method B Following Item Responses

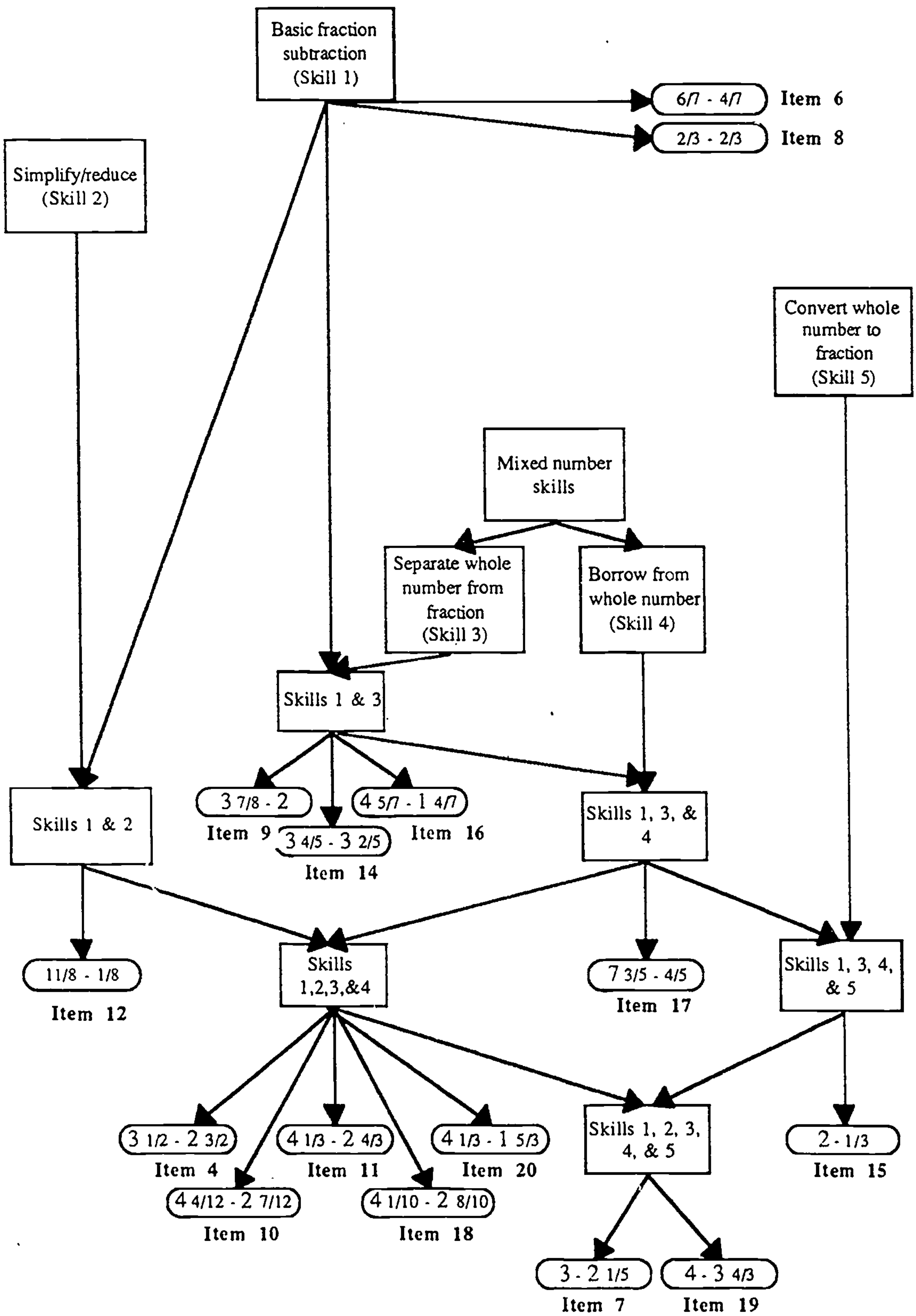


Figure 9

A Reduced Inference Network for Method B

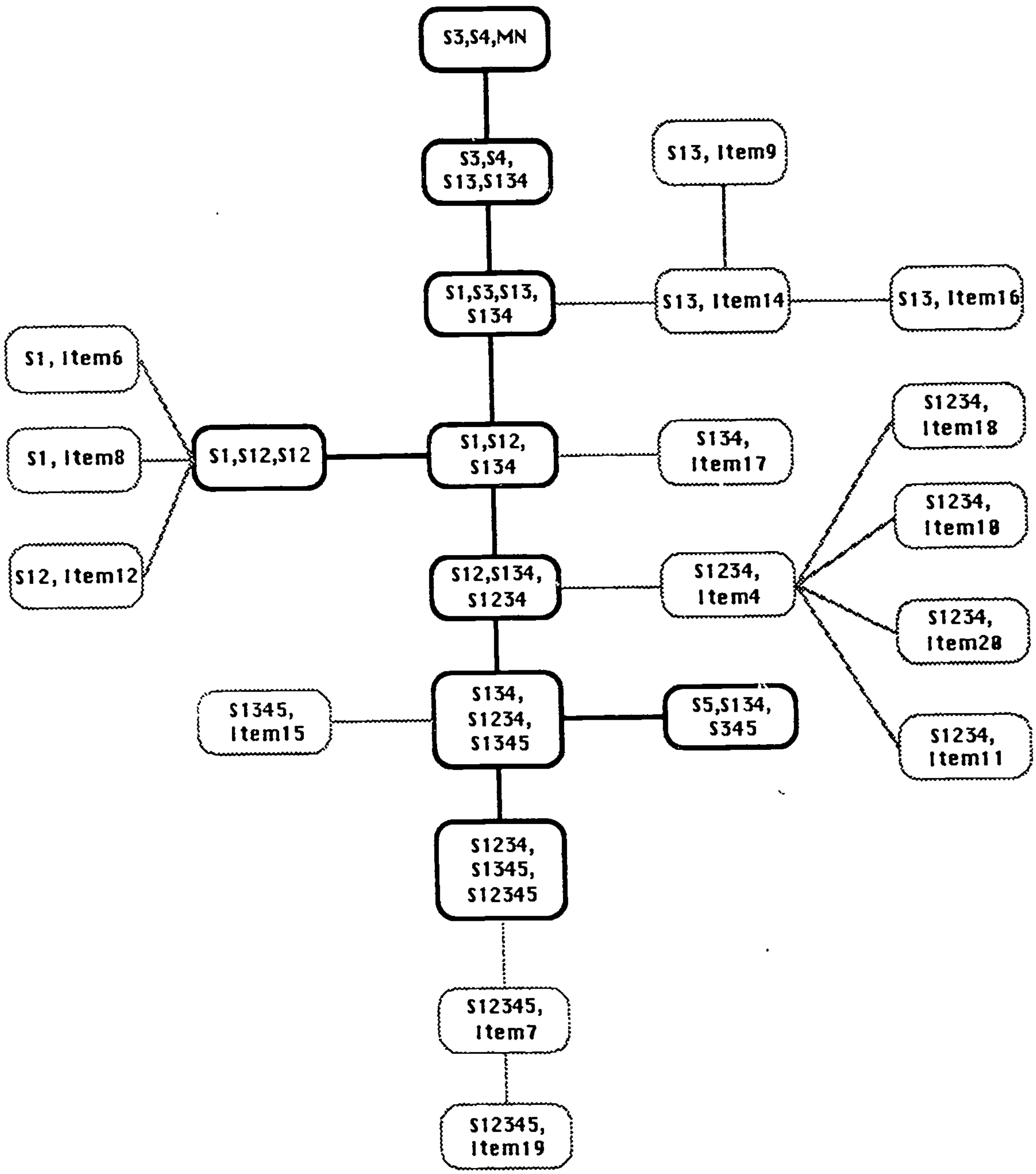


Figure 10

Join Tree for Network without Empirical Connections Among Skills

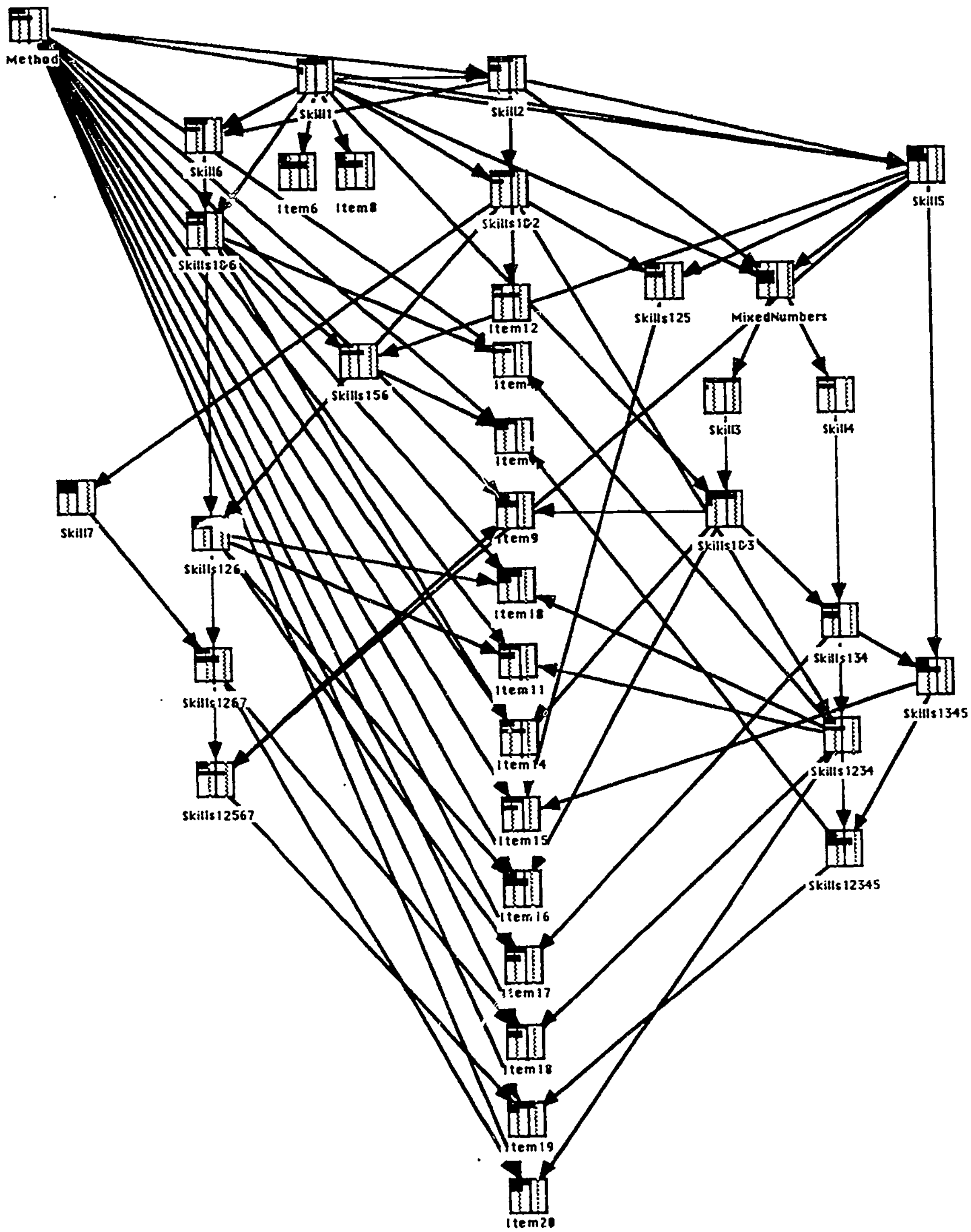


Figure 11
Prior Probabilities in Inference Network for Both Methods Combined

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Gregory Anrig
Educational Testing Service
Mail Stop 14-C
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr Isaac I Bejar
Educational Testing Service
Maii Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director
Life and Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
School of Education
Tel Aviv University
Ramat-Aviv 69978 ISRAEL

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Monterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Paul R Chatelier
PERCEPTRONICS
1911 North Ft Myer Drive
Suite 1100
Arlington VA 22209

Dr Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr William Crano
Department of Psychology
Texas A&M University
College Station TX 77843

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Meas Stat and Eval
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Dr Sharon Derry
Florida State University
Department of Psychology
Tallahassee FL 32306

Hei-Ki Dong
BELLCORE
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Educational Testing Service
Mail Stop 07-E
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd
Suite 550
Rockville MD 20850-3238

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Michael Kaplan
Office of Basic Research
US Army Research Institute
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick
Dept of Mathematics Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Logicon Inc (Attn: Library)
Tactical & Training Systems Div
PO Box 85158
San Diego CA 92138-5158

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr George B. Macready
Dept of Meas Stat & Eval
College of Education
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA 22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE 19716

Acad Prog & Research Branch
Naval Tech Training Command
Code N-62
NAS Memphis (75)
Millington TN 30854

Dr W Alan Nicewander
American College Testing
2201 North Dodge Street
. O Box 168
Iowa City IA 52243

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Cntr for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC 20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660
(6 Copies)

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-O1JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
NASA Ames Research Center
Mail Stop 239-1
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
American Council on Education
GED Testing Service Suite 20
One Dupont Circle NW
Washington DC 20036

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA 94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Dept of Industrial Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
WEST GERMANY

Dr David Wiley
Sch of Educ and Social Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Dept of Educational Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550